DOCUMENT RESUME

ED 434 153 TM 030 117

AUTHOR Tambini, Robert F.
TITLE An Analysis of the 1992 New Jersey Grade 8 Early Warning Test.
PUB DATE 1999-00-00
NOTE 10p.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Grade 8; *Junior High School Students; Junior High Schools; Standardized Tests; Test Construction; Test Content; *Test Use; *Testing Problems
IDENTIFIERS *Early Warning Test NJ

ABSTRACT

The quality and the effectiveness of the 1992 New Jersey Grade 8 Early Warning Test (NJEWT) are assessed. Standardized tests possess clear advantages for educators, especially in the case of administration and scoring, but there are clear disadvantages as well, including the possibility of bias. Four criteria are applied to the NJEWT: adequacy, impact, reliability, and validity. The writing test of the NJEWT appears to test reading comprehension more than writing, making the validity of information collected highly suspect. The quality of information provided by the reading test is also questionable, and the essays in this section are not well suited to the task of assessing student reading ability. The mathematics section also suffers some serious shortcomings. It uses word problems and requires students to explain in writing the reasoning they used to arrive at their answers, making it as much a test of language facility as of mathematics. Global deficiencies of the NJEWT include: (1) confusing structure; (2) confusing directions; and (3) confusing use of capital letters in the test items. The 1992 NJEWT has only limited value as an assessment tool. It is not possible to evaluate the reliability of the test in the absence of data from various test administrations across sample groups. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

Robert F. Tambini
Dept. of International Studies
Centenary College, Hackettstown, NJ

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Robert Tambini
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 434 153

1

An Analysis of the 1992 New Jersey Grade 8 Early Warning Test

Background

In this paper, I will endeavor to assess both the quality and effectiveness of the 1992 New Jersey Grade 8 Early Warning Test. This effort will begin with a brief discussion of the advantages and disadvantages of standardized tests as a means of assessment, followed by a description of the criteria selected for use in this analysis and the rationale for their selection. The test itself will then be discussed within the context of a rubric which has been specifically designed for the purpose of this analysis, and which employs the above-mentioned criteria. Finally, the conclusions drawn regarding the NJEWT will be generalized in a critique of standardized tests in general.

The use of standardized tests began in the nineteenth century, though this type of testing did not become widespread in the United States until the early part of the twentieth century. Initially, the primary focus of standardized testing was the measure of general intelligence, or IQ. However, by the 1950's, tests were being developed to assess student level and progress in an effort to improve the quality and effectiveness of educational services delivered at that time. Today, these tests are used in a great variety of ways, though their validity and

TM030117

BEST COPY AVAILABLE

reliability remain a source of discussion and controversy.

Advantages & Disadvantages

Clearly, standardized tests possess advantages for educators. For the most part, these advantages lay in the mechanics of the tests, as the validity of the instruments depends heavily on the form, manner, content, and administration of the tests themselves. Among these advantages are: objectivity – though the question of validity may be raised in this connection; ease of scoring – be it by machine or by professional interraters using scoring rubrics; ease of administration – as seen in the uniform delivery of tests via proctors; ease of information collection on a large scale – whether local, state-wide, or national; and, the ease with which scoring information gathered might be compared across sample groups.

However, there also appear to be distinct disadvantages in the use of standardized tests. Factors that could potentially impact the validity of the tests are the primary cause of concern among educators, parents, and students alike, as the results of these tests often influence the development of educational programs. Among these disadvantages are: the possible influence of bias – be it cultural, socio-economic, ethnic, gender (sexual), racial, or a combination of the above; the difficulty of assessing overall student abilities – standardized tests, traditionally, do not allow for performance-based assessment; the limitations placed on student creativity; the length of the tests and lack of variety in the

questions; and, the loss of class time as some teachers spend days or weeks "studying to" or preparing students for the tests.

Criteria & Analysis

In this analysis of the 1992 New Jersey Grade 8 Early Warning Test (NJEWT), I have chosen four criteria, assembled into a scoring rubric identifying those points under each criterion which might be met to a relatively greater or lesser degree, by which I will attempt to assess the test in question. These criteria were selected as I believe them to be relevant measures regarding the quality of the test as an assessment tool, and are: Adequacy - the quality of the procedures and manner of presentation; Impact - success of the instrument in meeting its stated goals; Reliability - consistency and repeatability of outcomes across time-frames and sample groups; and, Validity - alignment of the assessment tool with the objectives of the assessment.

The 1992 New Jersey Grade 8 Early Warning Test is divided into three sections: writing, reading, and mathematics. Each of these sections is of the same approximate length as the others, and the majority of questions are in multiple choice format. Interestingly, the writing section requires students to compose only a single essay, the rest of the test items in that section being primarily multiple choice questions; the reading section, on the other hand, contains numerous opportunities for students to display their writing ability, and the math section allows for written explanations of the ways in which students have

arrived at their answers. Based on these observations, it appears that the first thing which must be considered in analyzing the NJEWT is the question of validity - whether or not the test actually tests what it is intended to test.

In fact, insofar as the writing section primarily consists of a series of letters which students are to read, correcting perceived deficiencies by selecting from a list of possible alternatives, it seems that this section is in reality more a test of reading comprehension than it is of writing. In this section, following the initial writing prompt, students do not do any actual writing; they do, however, make choices based on their understanding of the statement in question as to how it might be improved or made more clear. Fixing someone else's sentence by making a selection from a predetermined list is not writing, but it does require that students have an understanding of the meaning of the sentence that they are being asked to improve. Therefore, as it appears clear that this section tests reading ability more than it does writing skill, the validity of any information gathered through the use of this test regarding writing must be considered highly suspect.

The quality of information provided by the reading section of the NJEWT must also be brought into question, though to a lesser degree than that of the writing section. This section is structured as a series of written pieces representing a variety of genres (fiction, non-fiction, epistolary), each followed by a number of multiple choice questions intended to assess levels of student comprehension and retention regarding that particular piece. While multiple

choice questions are an appropriate means by which to measure these abilities, the design of the test falters in that an essay component is included after each of the multiple choice sections.

Essays, by their very nature, provide students with the opportunity to showcase their writing skills. Placed within the context of the writing section of the NJEWT, essays are a potentially effective means by which the quality of student writing might be assessed. However, the essays contained in the reading section of the NJEWT, scored using the New Jersey Registered Holistic Scoring Rubric, are wholly unsuited to the task of assessing student reading ability. Because the inclusion of an essay component indicates that skills unrelated to reading are being assessed, the validity of any results collected concerning this section must be viewed as questionable.

The final section of the 1992 New Jersey Grade 8 Early Warning Test attempts to assess student competence in mathematics. Similar to the sections discussed above, the math section also suffers from some very serious shortcomings regarding the question of validity. Let me begin my description of the difficulties contained in this section with an example: As an English as a Second Language teacher who deals primarily with students from Japan, I have often observed the obstacles (for the most part language related) that my students must overcome if they are to succeed in college in the United States. However, of all the challenges that my students face, specifically those which they encounter in their academic lives, there is one subject area that they approach without fear or apprehension:

mathematics. This is because in a math class their relative lack of English ability does not present the difficulty that it does elsewhere, as the symbols and functions of mathematics are (insofar as this is possible) universally understood and accepted. Essentially, when it comes to the language of math, my international students are native speakers.

And therein lies the rub - Because the math section of the NJEWT solely employs word problems, and requires students to explain in writing the reasoning which they used to arrive at their answers, it is in fact less a test of mathematics than it is a test of reading comprehension and writing skill. Certainly, there are problems the solutions to which need be arrived at mathematically; however, students for whom language is problematic are at a distinct disadvantage when taking this test, especially should they be compared to students who are more facile in their manipulation and understanding of language-based items. Therefore, it is clear that the mathematics portion of the NJEWT assesses not only mathematical competence but language facility as well. As any outcomes recorded as a result of the administration of this test are unquestionably influenced by factors which lie without the bounds of the intended assessment, it is impossible to view those outcomes as possessing validity.

Regarding the NJEWT as a whole, there are certain global deficiencies affecting the validity of recorded outcomes which also might be pointed out and brought under discussion.

1) The overall structure of the test is confusing and could easily lead students to commit errors which, given a more clearly delineated presentation, they would not make otherwise. For example, consider the writing section following the initial writing prompt: The section begins with three correspondence, followed by a movie review, a biographical essay, and a short story, about which students are to read and make judgments. While the variety of genres represented is commendable, the main difficulty lay in the fact that the physical presentation of these items (more simply, their layout on the printed page) does nothing to distinguish one from the other. It is reasonable to believe that students, fully aware of the time limits under which they are working, might assume that this similarity of presentation indicated a congruity of expectations. Based on this assumption, it is likely that students would approach items four through six in the same manner as those that precede them.

2) The directions which are to be read by the proctor prior to each part of the three sections of the NJEWT are quite long and present a potential source of confusion for students. As an example, consider the explanation of the time allowed for part one of the reading section:

> You will have a total of 30 minutes to complete part 1... 20 minutes to read the story and answer the multiple choice questions, and 10 minutes to respond to the open-ended question and complete Part 1... I will keep track of the 30 minutes available... Work until you reach the end of the multiple choice questions. Do NOT go on to the open-ended question until you receive further directions.

Presented with directions such as these, students have every right to be confused. Thirty minutes are available, and the proctor will monitor the time; but, the last ten of those minutes seem to have been separated out somehow. Or have they? Are the thirty minutes continuous or aggregate? What are students supposed to do at the twenty-minute mark? Where the directions are unclear, so too must be students' understanding of what is actually expected of them.

3) The final point that I will make here (though, were it not for the limited space allowed for this current analysis, there are others to be made) regards the test items employed in the 1992 NJEWT. As a means of emphasizing certain key words in a large number of the questions, the test writers have chosen to print these words using all capital letters. The fundamental problem with this strategy lay not in the use of the capital letters (though this is a relatively unusual approach to emphasizing items of import), but rather in the fact that the reason for the capitalization of these words is in no way articulated to the students. This being the case, it is inevitable that, for some students at least, the apparently random use of capital letters within the text of the test items will lead to confusion and misunderstanding.

Conclusion

As a result of the foregoing analysis, it has become clear that the 1992 New

Jersey Grade 8 Early Warning Test possesses only limited value as an assessment tool. The validity of any outcomes recorded as a result of the administration of this instrument must be regarded as questionable at best. Similarly, in the areas of adequacy and impact, the test has also been found to be lacking. However, without further data comparing the results of various test administrations across sample groups, it is impossible to evaluate the reliability of the NJEWT, and for that reason the question of reliability has not been addressed in this study. Placed in the context of a rubric (Appendix A) specifically designed for the purpose of evaluating the quality and effectiveness of this test, the NJEWT scored as follows (based on a 4-point scale): Adequacy, 3; Impact, 2; Reliability, - [not addressed]; Validity, 2.

Any attempt to generalize the results of this analysis to include other standardized tests would, I believe, prove untenable. Well designed standardized tests, appropriately employed in the assessment of skills and abilities which lend themselves to this type of measure, serve an important function in the gathering and evaluation of information. Clearly, standardized testing provides an effective means by which writing, reading, and mathematics skills might be assessed; however, the 1992 New Jersey Grade 8 Early Warning Test suffers too many design flaws to be viewed either as an effective assessment instrument or as representative of the quality or value of standardized tests in general.